

Les gardiens du bon usage

Étude critique de « *Philosophical Foundations of Neuroscience* », de P. M. S. Hacker

et M. R. Bennett

Paru dans *Philosophiques*, vol. 34, no. 1, p. 183-200.

PIERRE POIRIER et NICOLAS PAYETTE

Université du Québec à Montréal

pierre.poirier@uqam.ca

nicolas.payette@internet.uqam.ca

1. Introduction

L'école philosophique qui s'est épanouie à l'Université Oxford entre les années 1930 et 1970 sous l'impulsion de Wittgenstein, Ryle, Austin, Strawson et Malcolm, entre autres, et que l'on connaît aujourd'hui sous le nom de « philosophie du langage ordinaire », ne compte plus beaucoup d'adeptes, et encore moins de pratiquants. Dans l'histoire de la philosophie analytique, on la voit souvent comme la première des deux réactions à la philosophie, quelquefois dite « du langage idéal », défendue par exemple par Russell, le premier Wittgenstein et Carnap, réaction qui s'est étiolée au contact de la seconde réaction à la philosophie du langage idéal, soit le mariage d'analyse conceptuelle, de naturalisme et de pragmatisme que l'on retrouve, sous diverses formes, chez Quine, Putnam et Rorty. La disparition de la philosophie du langage ordinaire de l'avant-scène philosophique ne signifie pas toutefois qu'elle soit morte et enterrée, comme en témoigne *Philosophical Foundations of Neuroscience*, l'ouvrage de Max R. Bennett et de Peter M. S. Hacker (B. et H. ci-dessous). On se souviendra que, selon cette école, comprendre la signification d'une expression linguistique, c'est suivre les règles qui guident son usage, et les philosophes créent une variété de faux problèmes lorsqu'ils emploient les mots hors du contexte de leur usage ordinaire. Dans l'ouvrage qui nous intéresse, ce ne sont toutefois pas les philosophes qu'on retrouve au banc des accusés, mais certains psychologues et neuroscientifiques, notamment ceux qui se réclament des *neurosciences cognitives*. Page après page, B. et H. y montrent patiemment, en multipliant les pièces à conviction, et ce non sans quelques longueurs, que les chercheurs et théoriciens des

neurosciences cognitives n'utilisent pas le vocabulaire psychologique (connaître, se rappeler, se représenter, emmagasiner de l'information, etc.) comme le prescrit l'usage ordinaire. Nous ne contesterons pas ici ce point puisque nous croyons que les auteurs l'ont amplement établi. Mais la question qui se pose – nous ne l'aborderons pas ici parce qu'elle nous amènerait sur le terrain d'une enquête empirique sur les neurosciences –, est de savoir si, dans ce cas précis, de réels faux problèmes découlent des usages non standards des neuroscientifiques ou si ceux-là n'ont d'autre effet que de choquer la bienséance des puristes de l'usage linguistique. Selon B. et H., tous ces usages non standards des chercheurs en neurosciences découlent d'une même source, soit le fait qu'ils attribuent au cerveau, ou à certains de ses processus, ou à des parties du cerveau, ou à certains de leurs processus, des propriétés que l'usage ordinaire attribue aux agents psychologiques (personnes, organismes, etc.). Sur le plan logique, cette déviation de l'usage ordinaire aurait pour source un sophisme méréologique de division, sophisme qui trouverait son origine chez Descartes. Puisque le sophisme méréologique est le genre de faute logique contre laquelle sont mis en garde les étudiants d'un cours de pensée critique, les neuroscientifiques rejeteront évidemment cette accusation, et les auteurs leur prêteront la parole pour se défendre, identifiant dans les écrits de ce domaine des penseurs qui offrent diverses défenses des positions neurologiques. Évidemment, B. et H. répondent à ces objections des neuroscientifiques, mais sans leur offrir un droit de réplique. C'est ce que nous ferons ici mais, avant d'entrer en mode polémique, nous allons d'abord présenter, à la section 2 ci-dessous, la nature et l'origine de ce qu'ils nomment le principe méréologique en neurosciences, puis nous illustrerons, section 3, leur critique en tant qu'elle s'applique à l'étude de la mémoire en neurosciences.

2. Le principe méréologique en neurosciences

Pour B. et H., le *principe méréologique en neurosciences* est un principe logique général qui circonscrit la classe des prédicats qu'on peut de bon droit attribuer aux structures neurologiques. Ce principe stipule simplement que l'attribution de prédicats psychologiques à des structures neurologiques est dénuée de sens. Chez les auteurs, l'attribut « dénué de sens » doit être compris comme le voulaient les positivistes

logiques : un énoncé attribuant une propriété psychologique à une structure neurologique n'a *aucune signification cognitive*, c'est-à-dire aucune valeur de vérité. Il n'est pas faux (ni vrai) mais simplement sans signification. Ainsi, pour prendre un exemple cité par les auteurs (p. 69), lorsque Antonio Damasio énonce, dans *L'Erreur de Descartes* (1997), que « nos cerveaux peuvent prendre de bonnes décisions en quelques secondes... », il dit quelque chose qui, à strictement parler, ne possède pas de valeur de vérité. Comme le rappellent les auteurs, Wittgenstein défend un principe semblable lorsqu'il soutient que « Seul d'un être humain vivant et de ce qui lui ressemble (dans son comportement) on peut dire : il a des sensations ; il voit ; est aveugle ; entend ; est sourd ; est conscient ou inconscient » (Wittgenstein, 1961, § 281).

Les neuroscientifiques et les psychologues produiraient ainsi des énoncés sans signification cognitive qui font fi du principe méréologique parce qu'ils commettraient un *sophisme méréologique* en attribuant à la partie (le cerveau ou ses régions, composants, etc.) des propriétés normalement attribuées aux êtres humains ou aux animaux, c'est-à-dire à des organismes en tant qu'ils manifestent des comportements. Rappelons que commettre un sophisme méréologique, c'est faire l'erreur de confondre le tout et la partie, et cette confusion peut être de deux types : a) attribuer à la partie une propriété du tout (sophisme méréologique de division) ; ou b) attribuer au tout une propriété de la partie (sophisme méréologique de composition). Le sophisme méréologique est une bête erreur de raisonnement. Il serait à tout le moins surprenant que *tous* les neuroscientifiques d'une époque commettent cette faute de débutant. Cette allégation exige une explication, et B. et H. se penchent sur l'histoire des neurosciences pour nous en offrir une.

Incidemment, c'est Descartes qui porte l'opprobre, ayant mis au rancart les neurosciences fondées sur la doctrine aristotélicienne, pourtant beaucoup plus saine sur le plan méréologique. Lorsque Aristote parlait des capacités psychologiques, il les attribuait à la « psyché » (ou « âme », *anima*), conçue d'emblée comme l'ensemble des fonctions essentielles d'une créature. Pas de sophisme méréologique ici, puisque la psyché n'est pas une partie du corps, mais bien la « forme » de celui-ci. Selon l'exemple d'Aristote, on ne peut pas plus séparer la psyché du corps qu'on ne peut séparer la capacité d'une hache à couper du bois du manche et de la lame de cette hache. Aristote distingue trois sortes de

psyché, chacune porteuse d'un ensemble de propriétés différentes : la croissance, la nutrition et la reproduction sont le propre de l'« âme nutritive » ; la perception, le désir et la locomotion appartiennent à l'« âme sensitive » et finalement, le raisonnement et la volition nécessitent une « âme rationnelle ». Cette conception de la psyché est conservée chez Galen (130-200 av. J.-C.) qui, lorsqu'il découvre que les nerfs jouent le rôle de conduits pour les sensations et les commandes motrices, postule lui aussi une « âme sensorielle » et une « âme motrice » qui constituent deux principes d'activité distincts. Ce genre de conception aristotélicienne, quoiqu'il ne fasse pas l'unanimité (Nemesius par exemple, le père de la doctrine ventriculaire, adhérerait plutôt à une forme de dualisme néoplatonicien, métempsycose comprise), accompagnera le développement des neurosciences jusque chez Jean Fernel (1495-1558), auquel B. et H. attribuent par ailleurs le premier traitement formel de la physiologie. Fernel hérite toutefois des « confusions » nées des tentatives scolastiques de réconcilier la doctrine chrétienne de l'immortalité de l'âme avec la psyché aristotélicienne. Descartes, par contre, ne s'embarrassera pas de tels compromis : la coupure entre le corps et l'âme est chez lui bien nette.

B. et H. relèvent quatre marques de la transformation radicale que Descartes fait subir au concept d'âme. La première est qu'il n'y a pas chez Descartes d'âme autre que l'esprit. Les fonctions de l'âme nutritive et la plupart de celles de l'âme sensitive d'Aristote sont reléguées au statut de fonctions corporelles qui doivent être conçues en termes entièrement mécanistes. Cette restriction s'accompagne d'une extension dans une autre direction, qui constitue la seconde marque de la transformation cartésienne : l'esprit est identifié à la pensée, elle-même définie en termes d'expérience consciente. La perception sensorielle et l'imagination, propriétés de l'âme sensitive chez Aristote, figurent dans l'esprit cartésien à côté du raisonnement et de la volonté. Le troisième changement apporté par Descartes est bien connu : c'est l'adoption du dualisme radical. Le corps et l'esprit ne sont plus deux aspects d'une seule et même chose, mais bien deux substances distinctes, dont les interactions seront désormais l'objet des neurosciences. Finalement, chacune des deux substances cartésiennes possède une seule propriété essentielle : la pensée pour l'esprit (la *res cogitans*), et l'extension pour la matière (la *res extensa*) et donc pour le corps. Cette conception du corps en tant que substance étendue

évacue la téléologie chère à ses prédécesseurs et amène la physiologie et les neurosciences sur le terrain de l'analyse mécanique. C'est peut-être, admettent B. et H., la plus grande contribution de Descartes aux sciences biologiques.

La transformation cartésienne a toutefois pour conséquence malheureuse de plonger les neurosciences en pleine confusion méréologique. Cette confusion naît du problème auquel Descartes fait face en ce qui concerne l'union de l'âme et du corps. On se rappellera que, pour Descartes, c'est la glande pinéale qui est le lieu de l'interaction entre les deux substances : les événements sensoriels affecteraient causalement cette glande, ce qui permettrait à *l'âme* de percevoir des images et de former des représentations. Ce faisant, Descartes tombe dans le piège consistant à attribuer à une partie (l'âme), ce qui ne peut logiquement, selon B. et H., n'être attribué qu'au tout (la personne). Ce faux pas cartésien serait un moindre mal s'il n'avait eu la postérité qu'on lui connaît. B. et H. retracent cette postérité dans le détail, ce que nous ne ferons pas ici. Qu'il suffise de dire que le problème de l'union de l'âme et du corps continue à faire couler beaucoup d'encre et que les successeurs de Descartes, de Thomas Willis (1621-75) à John Eccles (1903-97), persistent à la fois dans un dualisme plus ou moins cartésien et dans la confusion méréologique. Et si la génération actuelle de neuroscientifiques et de psychologues ont majoritairement délaissé le dualisme au profit d'un monisme matérialiste, ils n'échappent pas pour autant à l'erreur méréologique puisque les caractéristiques auparavant attribuées erronément à l'âme sont maintenant attribuées, tout aussi erronément, au cerveau, ce que B. et H. qualifient d'« adhésion irréfléchie à une forme mutante de cartésianisme » (p. 72).

B. et H. ne manquent pas d'exemples pour appuyer ce dernier constat. Outre celui de Damasio cité plus haut (le cerveau qui décide), on a le cerveau qui « croit », « interprète » et « devine » (Crick, 1995, pp. 30, 32n., 57) ; qui « catégorise », qui « distingue » et qui « manipule des règles » (Edelman, 1994, pp. 109n., 130) ; qui « connaît », qui « raisonne » et qui « construit des hypothèses » (Blakemore, 1977, p. 91) ; qui « pose des questions » et « cherche des réponses » (Young, 1978, p. 119) ; et la liste s'étend à maintes autres facultés (voir pp. 68-70). On a même chez Blakemore (*ibid.*) des neurones « intelligents » qui « estiment des probabilités ». Bref, ces façons de

parler sont omniprésentes dans le discours neuroscientifique contemporain. Un discours qui, selon B. et H., est gravement malade et a besoin de la philosophie à son chevet pour lui administrer une bonne dose de clarification conceptuelle. À défaut de quoi, il est condamné à ne pas avoir de sens et à abandonner, conséquemment, toute prétention de vérité.

Évidemment, la valeur du diagnostic dépend de celle du principe méréologique : assurément les neuroscientifiques commettent un sophisme méréologique lorsqu'ils attribuent une propriété psychologique à une structure neurologique s'il est vrai qu'il n'y a aucun sens à attribuer une propriété psychologique à autre chose qu'un organisme complet (comme un humain). Or on peut se douter que les neuroscientifiques n'accepteraient pas ce principe et que, s'ils avaient la chance de répondre, ils mettraient en doute sa valeur, ou à tout le moins sa préséance absolue. À la fin de ce commentaire, nous tenterons d'ébaucher le genre d'argument qu'ils devraient selon nous opposer au principe méréologique. Avant de formuler ces objections, nous présenterons toutefois les raisons invoquées par B. et H. en faveur du principe et nous illustrerons en détail son application à l'attribution d'une capacité psychologique : la mémoire.

Selon les auteurs, nous reconnaissons l'applicabilité d'un prédicat psychologique sur la base de critères d'évidence (*criterial grounds*), principalement des évidences comportementales, et non sur la base d'inférences inductives basées sur le comportement observé. Nous ne voyons pas une personne gémir puis en concluons par induction qu'elle souffre (probablement) : nous *voyons* qu'elle souffre, et son gémissement sert d'évidence pour justifier l'attribution de souffrance à la personne. Au besoin, par exemple lorsqu'on nous le demande, nous nous justifions en citant les évidences comportementales qui servent de critère à cette attribution. Selon B. et H., ces évidences comportementales constituent en partie la signification des prédicats psychologiques. Cela ne veut pas dire qu'il soit impossible d'établir des corrélations entre des états psychologiques et des événements neurophysiologiques, mais que de telles corrélations sont logiquement subordonnées à nos critères habituels d'attribution des prédicats psychologiques. Qu'il se produise telle ou telle chose dans le cerveau lorsqu'une créature souffre est une découverte empirique qui ne pourrait avoir lieu si nous n'avions pas déjà les critères d'attribution

inhérents au concept de douleur. En cas de conflit avec les observations neurophysiologiques, tranchent B. et H., ce sont les critères comportementaux qui priment.

Puisque le cerveau ne manifeste pas de comportements (il ne gémit pas, etc.), il ne satisfait pas les critères d'application des prédicats psychologiques. C'est pourquoi il est épistémiquement erroné d'appliquer des prédicats psychologiques au cerveau ; et aussi pourquoi les phrases résultant de telles applications sont sans signification. Lorsque nous attribuons, erronément, des prédicats psychologiques au cerveau, nous le faisons à la suite d'une inférence inductive. Nous voyons que la personne a appris une information, nous voyons que telle activité cérébrale chez cette personne est corrélée à cet apprentissage, et nous en concluons, erronément, que le cerveau a appris. Tout ce que nous pouvons vraiment conclure, c'est que l'activité en question est corrélée avec l'apprentissage de la personne.

Pourquoi alors les neuroscientifiques sont-ils tentés d'attribuer des propriétés psychologiques au cerveau ? Nous avons déjà évoqué les origines cartésiennes attribuées par B. et H. à ces façons de parler, mais les auteurs poussent leur analyse encore un peu plus loin et révèlent un mal encore plus profond. C'est que les neuroscientifiques souscriraient aussi à certaines thèses que le cartésianisme partage avec l'empirisme classique. Ces thèses seraient fondées sur des mauvaises conceptions du psychologique, notamment quant aux propriétés psychologiques (qui seraient « intérieures » et « mentales »), quant à l'introspection (qui serait une forme de perception), quant à « l'accès privilégié » et « l'accès direct » à l'expérience, quant au caractère privé et subjectif de l'expérience et enfin quant à la signification des prédicats psychologiques. Après avoir montré par diverses citations que les neuroscientifiques souscrivent bel et bien à ces thèses, les auteurs expliquent comment elles sont fondées sur de mauvaises conceptions et comment, le cas échéant, on devrait plutôt comprendre les éléments en jeu.

Ces mauvaises conceptions du psychologique sous-tendent entre autres une conception erronée de la sémantique des prédicats psychologiques, à savoir que ceux-ci nomment des états internes comme le nom « pomme » désigne un certain fruit. Les termes comme « pomme » sont définis ostensivement (ce fruit < INSÉRER UNE POMME ICI >

s'appelle « pomme ») et la conception cartésienne/empiriste de l'esprit nous laisse croire que les termes psychologiques sont également définis ostensivement (cette sensation < INSÉRER UNE DOULEUR ICI > se nomme « douleur »). Logiquement, les prédicats psychologiques comme « avoir mal », « désirer que », « avoir l'intention de » et « croire que » ne sont pas des noms qui acquièrent leur signification par définition ostensive (les auteurs justifient cette thèse par le fameux argument du langage privé de Wittgenstein). Ces prédicats n'ont même pas une sémantique unique (croire qu'ils en possèdent une est une autre conséquence de la conception cartésienne/empiriste) : « avoir mal » et « désirer que » sont des extensions linguistiques de comportements naturels (respectivement de comportements de douleur et de comportements conatifs) ; « croire que » (aussi « savoir que ») est un opérateur linguistique pour exprimer notre accord avec une certaine proposition ou nos fondements pour la soutenir ; « avoir l'intention de » est aussi un opérateur linguistique, dans ce cas pour donner un fondement à une prédiction comportementale (si quelqu'un dit : « J'ai l'intention de partir », alors il nous donne un fondement pour prédire qu'il partira). Il est facile de voir comment la conception cartésienne/empiriste de la sémantique des prédicats psychologiques peut mener à l'attribution de ces prédicats au cerveau : si l'on considère les termes psychologiques comme étant des noms pour des entités intérieures, accessibles à la conscience, et que ces entités sont localisées dans le cerveau, il est tout naturel de franchir le pas faisant du cerveau le possesseur de ces entités. Ce pas, B. et H. montrent bien que les neuroscientifiques et les psychologues le franchissent allégrement. Nous verrons plus loin ce qu'ils ont à dire pour leur défense, mais pas avant d'avoir donné un exemple du genre de critique à laquelle les auteurs consacrent la majeure partie de leur ouvrage.

3. Une critique linguistique des théories neurologiques de la mémoire

B. et H. appliquent leur analyse conceptuelle à une multitude d'habiletés psychologiques (par exemple avoir des sensations et perceptions), de capacités cognitives (connaître et mémoriser) et cogitatives (croire, penser, imaginer). L'analyse est également appliquée aux émotions, à la volition et à la conscience. Il serait évidemment trop long de reprendre chacune de ces analyses ici, et le résultat serait beaucoup trop superficiel pour être

satisfaisant. Nous proposons plutôt de reprendre une de ces analyses en détail, celle qui concerne la mémoire, c'est-à-dire la « faculté de *retenir* la connaissance acquise » (p. 154). Dans cette section, nous tenterons surtout de donner la saveur des analyses des auteurs, en gardant autant que possible nos réponses ou nos critiques pour la prochaine section.

Selon B. et H., la mémoire est une capacité cognitive des organismes, en particulier des êtres humains. Pour un être humain, se remémorer, c'est ramener à son esprit une connaissance qu'il a retenue (p. 155). Et bien que, pour ce faire, il doive en être venu à connaître ce dont il se souvient, il ne s'ensuit pas que la mémoire concerne spécifiquement le passé. On peut se souvenir de faits présents (comme lorsque je me rappelle que la lampe est allumée dans la chambre), futurs (comme lorsque je me rappelle que l'hiver arrivera autour du 21 décembre) ou intemporels (comme lorsque je me rappelle que l'hypoténuse d'un triangle rectangle est égale à la racine carrée de la somme du carré des deux autres côtés). Ils soulignent, suivant Malcolm (1963, pp. 203-221), que la mémoire, en tant que capacité cognitive d'un être humain, peut prendre trois formes : *factuelle* (ramener à l'esprit des connaissances factuelles, comme les trois précédents exemples), *expérientielle* (ramener à l'esprit des expériences, comme se rappeler avoir traversé l'Amérique en voiture) et *objectuelle* (ramener à l'esprit des objets, des personnes, comme se rappeler la maison de son enfance – à distinguer du fait de se rappeler que l'on avait une telle maison, ce qu'on exprime parfois en disant : « Oui, je me souviens bien que j'habitais cette maison pendant mon enfance, mais je ne me souviens plus de la maison elle-même »). Une dernière distinction linguistique est amenée par B. et H. avant de passer en mode critique. Il s'agit de la distinction entre *se souvenir que* (je me souviens que la neige arrive d'habitude en novembre) et *se souvenir comment* (je me souviens comment patiner sur de la glace). Comme avec la distinction plus connue de Ryle (*savoir que* et *savoir comment*), ces deux formes de souvenir ne sont pas mutuellement exclusives, puisque, dans bien des cas, se souvenir *comment*, c'est se souvenir *que* l'on fait une chose de telle et telle manière.

Le sujet grammatical de toutes ces capacités est un être humain et non, comme le supposent les neuroscientifiques, le cerveau ou le système nerveux, ou une partie de

ceux-ci. Parce que, selon B. et H., les seules normes dont nous disposons pour juger de l'application adéquate ou non des prédicats concernés sont de nature comportementale (nous voyons que Paul se souvient comment faire de la bicyclette lorsque nous le voyons en faire), et que le comportement en question est celui d'un être humain, il s'ensuit que les prédicats ne s'appliquent pas à des cerveaux ou des parties de cerveaux (le cortex, l'hippocampe, ou une autre structure). Comme nous l'avons dit précédemment, ce n'est pas qu'il soit faux de dire que le cerveau se souvient (ou apprend, ou emmagasine) : de telles phrases sont tout simplement dénuées de sens. Dans le cas particulier de la mémoire, les auteurs corrigent les erreurs des neuroscientifiques sur deux points en particulier : (i) dans leur traitement de la mémoire – déclarative et non déclarative –, et (ii) dans leur usage d'expressions comme l'*emmagasinement*, la *rétenion*, et les *traces mnésiques*. Nous reprenons ces critiques tour à tour.

La distinction mentionnée ci-dessus entre *se souvenir que* et *se souvenir comment* est associée dans la littérature neuroscientifique à la distinction entre la mémoire déclarative et la mémoire non déclarative. En neurosciences et en psychologie, la mémoire déclarative est la mémoire des faits : elle peut être exprimée linguistiquement, peut être vraie ou fausse, et permet d'emmagasiner des faits et des événements. Selon B. et H., bien qu'important, le travail des neuroscientifiques sur la mémoire déclarative est mal conçu lorsque compris comme impliquant l'emmagasinement d'informations ou la construction d'un modèle du monde. Ce sont les humains qui emmagasinent de l'information (par exemple, lorsqu'on emmagasine l'heure des marées dans un almanach) ou qui construisent des modèles du monde (ce que fait un maquettiste, par exemple). Selon les auteurs, on doit plutôt concevoir le travail des neuroscientifiques comme un travail concernant les changements neuronaux qui sont la cause de la rétention de l'information factuelle (déclarative) par l'individu. À part cette remarque au sujet de l'usage, qui est une application directe de leurs thèses plus générales sur le principe méréologique, les auteurs ne font aucune critique plus précise sur le traitement neurologique de la mémoire déclarative. Ils questionnent par ailleurs une thèse empirique qu'on rencontre chez certains auteurs, selon laquelle la mémoire déclarative est toujours associée à une expérience consciente (l'expérience de remémoration). Selon eux, et nous

sommes d'accord, il est faux de dire que la mémoire déclarative est *toujours* associée à une expérience consciente de remémoration : je peux me rappeler qu'on peut tourner à droite au feu rouge sans qu'aucune expérience consciente n'y soit associée. Mais si nous sommes ici d'accord avec eux, c'est que cette critique est de nature empirique (introspective) et que, là, ils ont vu juste.

Dans la littérature neuroscientifique, la mémoire non déclarative est comprise comme sous-tendant divers changements neurologiques permettant de s'adapter à l'environnement, que ce soit par sensibilisation, par habituation, par amorçage ou par conditionnement classique. Selon B. et H., la confusion des neuroscientifiques est ici plus profonde, car toutes ces capacités ne sont pas des formes de mémoire, laquelle est toujours *cognitive*, mais des formes non cognitives d'adaptation à l'environnement. Bien qu'il s'agisse dans tous ces cas de formes d'apprentissage, selon les auteurs, les apprentissages ne sont pas tous de l'acquisition de connaissances (donc de la mémorisation). **Nous sommes d'accord sur ce point mais remettons en doute son importance. L'idée d'inclure toutes les formes d'adaptation non cognitives à l'environnement, de l'amorçage au conditionnement classique, sous le vocable « mémoire » dans le contexte des théories dites des « systèmes de mémoire multiples », est récente (elle date des années 1980) et sans grandes conséquences conceptuelles ou empiriques. Imaginons que, convaincue par les propos des auteurs, la communauté se ravise et choisisse plutôt de parler de « systèmes d'adaptation ontogénique multiples », dont l'un est le système cognitif d'adaptation (la mémoire, qui était sous l'ancien système « la mémoire déclarative ») et l'autre regroupe les systèmes sous-cognitifs d'adaptation ontogénique (et incluent la sensibilisation, l'habituation, l'amorçage et le conditionnement classique). Qu'est-ce d'autre que des mots qui aurait changé ? Si la section n'est censée qu'exposer la critique de B. et H., ce commentaire n'appartient-il pas plutôt à la note de bas de page ? Sinon, ne faudrait-il pas modifier le commentaire du début selon lequel, dans cette section, B. et H. se contentent de présenter leur point de vue ?**

Les auteurs terminent le chapitre par une analyse de trois concepts associés à la mémoire : l'emmagasinement, la rétention et les traces mnésiques. L'erreur principale des

neuroscientifiques est de concevoir la mémoire comme de la connaissance emmagasinée (pire : emmagasinée dans le cerveau). Les auteurs retracent cette conception erronée dans l'histoire, d'Aristote et Platon jusqu'aux neuroscientifiques contemporains (LeDoux, Squire, Kandel, etc.) en passant par les empiristes classiques et les psychologues du XIX^e (James) et du XX^e siècle (Köhler). Ils analysent ensuite quatre idées douteuses impliquées dans cette conception de la mémoire comme emmagasinement d'informations ou de connaissances : (i) nous emmagasinons ce que nous percevons, (ii) ce que nous emmagasinons, une mémoire (*a memory*) ? un souvenir ? *représente* l'expérience originelle qui est mémorisée, (iii) la trace mnésique est déposée dans le cerveau par le biais de changements synaptiques et donc ceux-ci contiennent une représentation de l'expérience originelle ; (iv) et enfin la remémoration consiste en une recreation du patron originel d'activation des neurones.

Selon B. et H., une erreur conceptuelle importante causée par ces idées douteuses peut être de laisser croire aux neuroscientifiques que la mémoire n'est que mémoire du passé : comment le cerveau pourrait-il emmagasiner des traces mnésiques du présent ou du futur ? Or, comme nous l'avons souligné ci-dessus, on peut se souvenir de faits passés, actuels, futurs ou intemporels. Le cadre conceptuel erroné des neuroscientifiques, et notamment la proposition (ii) ci-dessus, les obligerait donc à penser que, dans ces cas, nous ne nous souvenons pas du présent (que la lampe est actuellement allumée dans la chambre) ou du futur (que l'hiver arrivera le 21 décembre) mais des événements passés associés (l'événement passé : j'ai allumé la lampe, ou le moment où j'ai appris que l'hiver arrivait le 21 décembre). Selon eux, cela est absurde, et ils ont certainement raison. Mais on peut se demander si les neuroscientifiques pensent vraiment ce genre de choses, notamment s'ils accepteraient bien cette proposition (ii). Ils pourraient répondre que, lorsqu'on se souvient d'un événement futur, c'est que s'active dans notre cerveau une trace faite dans le passé *mais à propos de, ou représentant un événement qui surviendra dans le futur*. Lorsqu'on se souvient d'un événement futur, on ne se souvient pas de la trace mais d'un élément de son contenu. En fait, on ne se souvient jamais de la trace : la trace est le moyen par lequel on se souvient. On pourrait aussi, si on le voulait, se souvenir de l'événement qui a laissé la trace (ou du moins, chercher à le faire). Ainsi,

une personne peut se rappeler qu'elle doit aller dîner avec un ami la semaine prochaine, et ce souvenir porte sur le futur. Mais elle pourra également se rappeler le moment où elle a accepté l'invitation de son ami, événement qui a laissé la trace, donc le contenu porte sur le futur. Les deux souvenirs sont indépendants, quoiqu'ils soient souvent, dans les faits, invoqués ensemble.

4. Entre anarchie et inertie sémantique

Nous avons promis plus haut que nous allions donner un droit de réplique aux neuroscientifiques, et c'est ce à quoi nous allons consacrer cette dernière section. Nous procéderons plus précisément de la façon suivante : nous exposerons d'abord quatre répliques possibles, formulées par B. et H. eux-mêmes, que les neuroscientifiques pourraient objecter au principe méréologique. Nous noterons au passage les réponses que les auteurs formulent contre ces répliques. Nous construirons ensuite notre propre argument en faveur du maintien d'une certaine latitude sémantique en neurosciences, c'est-à-dire une latitude certaine mais limitée, et ce afin de trouver un milieu entre l'inertie sémantique que nous entrevoyons comme conséquence des propos de B. et H. et l'anarchie sémantique qu'ils voient dans les écrits des neuroscientifiques. Les trois premières répliques sont présentées au chapitre 3, et la quatrième est présentée au chapitre 14. En ce qui concerne les trois premières répliques, nous ne respecterons pas l'ordre d'exposition utilisé par B. et H. Nous nous permettons cette réorganisation parce qu'il nous semble y avoir entre elles une certaine progression logique, que nous essaierons de faire ressortir dans le texte.

La première objection que nous allons considérer (la deuxième traitée par B. et H.) veut que les termes psychologiques utilisés par les neuroscientifiques soient des extensions, par analogie, des termes usuels. L'exemple classique d'une telle extension par analogie est celle du vocabulaire de l'hydrodynamique dans la théorie de l'électricité. Cette « analogie hydraulique » s'est non seulement révélée fructueuse lors du développement de la théorie de l'électricité, mais elle est encore aujourd'hui un outil pédagogique précieux pour l'enseignement de cette théorie. Le vocabulaire psychologique pourrait-il jouer la même fonction en neurosciences ? C'est la prétention

que B. et H. attribuent à Richard Gregory (1973, p. 51), prétention avec laquelle ils ne sont évidemment pas d'accord. Leur désaccord repose en partie sur le principe méréologique, lequel n'est pas transgressé par l'analogie entre l'électricité et un fluide ou entre un fil électrique et un tuyau. La théorie des fluides établit entre ses objets des relations qui ont bel et bien une parenté structurelle avec la théorie de l'électricité. Ces relations sont exprimées par des équations qui peuvent servir à justifier des énoncés tels que : « Le courant est inversement proportionnel à la résistance du fil conducteur *comme* le débit est inversement proportionnel au diamètre du tuyau ». Dans le meilleur des cas, une bonne analogie peut même servir de support inductif. L'analogie entre un humain qui pense et un cerveau qui pense, par contre, ne pourrait se prévaloir de la même correspondance structurelle puisque la pensée est attribuée au tout dans le premier cas et à la partie dans le second. De plus, l'analogie avec le vocabulaire psychologique ne partagerait pas selon B. et H. le caractère formel de l'analogie électro/hydrodynamique, et ne mettrait donc pas à la disposition des neurosciences l'appareillage qui serait nécessaire à sa justification. Au chapitre 13, qui porte sur le réductionnisme, B. et H. font grand cas du caractère non théorique du vocabulaire psychologique, lequel, selon eux, met celui-ci à l'abri de toute réduction interthéorique. La question du réductionnisme est assez périphérique dans leur ouvrage (et nous ne la traiterons pas ici), mais elle met en relief le statut particulier qu'ils accordent au vocabulaire psychologique : non seulement la science n'a pas droit de regard sur les règles qui le régissent, mais elle ne doit pas non plus oublier qu'il n'est pas soumis aux mêmes critères épistémologiques qu'elle. Les « fonctions [du vocabulaire psychologique] sont aussi diverses que celles du langage humain et ses ramifications tout aussi étendues que celle du phénomène de la vie humaine elle-même. » (p. 375). Puisqu'ils conçoivent en ces termes le rôle du vocabulaire psychologique, il n'est pas surprenant que B. et H. écartent la possibilité de l'utiliser comme support analogique pour penser les neurosciences. Comment, en effet, établir une analogie bien contrôlée avec un réseau conceptuel dont les ramifications sont aussi étendues ? Si les neuroscientifiques ne sont pas en mesure d'établir qu'il existe une analogie satisfaisante entre le vocabulaire psychologique ordinaire et celui des neurosciences, ils peuvent néanmoins tenter d'invoquer le recours à une figure

linguistique moins contraignante : la métaphore. C'est sur l'utilisation de celle-ci que repose l'objection suivante.

Cette seconde objection (la troisième traitée par B. et H.) que les neuroscientifiques pourraient opposer à leur argument est attribuée à Colin Blakemore (1990, pp. 257-283). Selon cette réplique, les neuroscientifiques savent très bien de quoi ils parlent lorsqu'ils expriment leurs découvertes. Les référents de leurs idées ou de leurs concepts sont alors bien clairs. Cependant, continue la réplique, la langue (anglaise en l'occurrence) ne possède pas les termes pour exprimer ces concepts ; ce qui n'est pas surprenant puisque c'est la première fois que des humains appréhendent le type de phénomènes qui occupent les neuroscientifiques. Pour exprimer ces concepts aux référents clairement identifiés, et ce malgré les carences de la langue anglaise, les neuroscientifiques doivent donc utiliser les termes usuels de manière métaphorique et figurative. Ainsi, lorsqu'ils parlent de « cartes » dans le cerveau, ils savent bien qu'il n'y a pas, dans le cerveau, un petit cartographe (avec des petites feuilles sur lesquelles tracer ses cartes et de petits crayons pour les tracer). Le terme « carte » exprime alors une idée bien précise : *le patron topographique de l'activité nerveuse observé dans diverses aires cérébrales*, ou encore *les dépendances topographiquement systématiques entre les variations de l'activité nerveuse et les variations d'une portion délimitée du champ perceptuel*. Le problème, selon B. et H., c'est qu'il est très difficile alors de ne pas se laisser emporter par l'usage métaphorique et ainsi faire des inférences illégitimes. D'ailleurs, ils illustrent leur propos au moyen d'une telle inférence illégitime faite par Blakemore lui-même lorsqu'il dit que ces cartes servent à l'« interprétation » du monde par le cerveau. Selon eux, ces métaphores sont autant de pelures de bananes (leur expression) que les neuroscientifiques lancent sur le chemin et qui peuvent les amener (eux-mêmes, d'autres neuroscientifiques ou le public) à glisser et à faire de mauvaises inférences. Cependant, pourrions-nous répondre, tout ce que B. et H. prouvent ici, c'est que les usages métaphoriques en génèrent d'autres, ce qui n'est pas un problème tant qu'ils ne nous ont pas montré que l'usage métaphorique des termes psychologiques amène de véritables fautes conceptuelles ou empiriques. Quoi qu'il en soit, les auteurs répondront finalement à cette réplique plus loin dans l'ouvrage, et de manière assez

cinglante : « Ce ne sont pas les carences de la langue anglaise qui sont responsables des confusions et des incohérences que l'on retrouve dans les écrits de tant de neuroscientifiques cognitifs – on doit plutôt celles-ci à des erreurs de compréhension, c'est-à-dire à la compréhension inadéquate des concepts et liens conceptuels en termes desquels ces scientifiques décrivent leurs découvertes et s'efforcent d'expliquer leurs implications pour notre compréhension de la psychologie humaine » (p. 379). Pour le prouver, ils invoquent leur re-description des résultats de la commissurotomie en termes méréologiquement adéquats, où ils remplacent des énoncés comme « l'hémisphère gauche ne sait pas ce que l'hémisphère droit a vu » par « la stimulation visuelle de l'hémisphère droit est déconnectée de l'hémisphère gauche » (p. 392). La description proposée par B. et H. montre qu'il est tout à fait possible de se passer des métaphores psychologiques, mais ce qu'elle gagne en rectitude linguistique, elle le perd en concision et en expressivité. Il serait sans doute intéressant de comparer les vertus pédagogiques de l'une et l'autre en les faisant lire à deux groupes de sujets différents et en comparant les résultats obtenus par ces sujets à un test leur demandant, par exemple, de prédire le comportement de patients ayant subi une commissurotomie. Mais de tels résultats empiriques (et les conséquences pragmatiques qui en découlent) n'auraient, bien sûr, aucun impact sur la philosophie telle que conçue par B. et H.. Soulignons aussi que chaque traduction réussie d'un usage métaphorique en termes non métaphoriques montre bien que les neuroscientifiques savent de quoi ils parlent lorsqu'ils utilisent leurs métaphores. La situation qui poserait problème serait celle où il est impossible de traduire un usage métaphorique en termes non métaphoriques. Au mieux, et en l'absence de telles démonstrations, nous croyons que les arguments de B. et H. pourraient simplement servir de mise en garde aux neuroscientifiques contre les possibles méfaits des inférences basées sur les métaphores. Mais si l'on considère malgré tout que les figures linguistiques qui tentent de préserver une communauté de sens entre l'usage scientifique et l'usage ordinaire échouent, les neuroscientifiques peuvent encore tenter de préserver l'usage de termes communs dans le discours scientifique en abandonnant la prétention d'en préserver le sens.

Alors que les deux objections précédentes plaidaient pour des usages non littéraux des termes psychologiques (l'un par analogie, l'autre par métaphore), la troisième objection que nous considérons ici (mais la première traitée par B. et H.) pose qu'un *nouveau* sens est créé par les usages des neuroscientifiques. Selon cette objection, que les auteurs attribuent à Ullman (1991, p. 314n.), les termes psychologiques utilisés par les neuroscientifiques sont des homonymes des prédicats du langage ordinaire, mais ce sont en fait des termes différents qui ont un sens technique distinct de celui de leurs homonymes psychologiques. Ainsi, les neuroscientifiques ne soutiendraient pas que le cerveau *pense*, *croit*, *infère*, *interprète*, etc., mais plutôt que le cerveau *pense**, *croit**, *infère**, *interprète**, etc., les termes affublés d'une étoile n'étant que des termes neuroscientifiques vaguement apparentés aux termes originaux. Puisque ce sont des termes différents, ils ne sont pas soumis aux conventions de l'usage ordinaire, mais bien à des règles d'utilisation théorique que les neuroscientifiques sont libres de définir comme bon leur semble. B. et H. voient deux problèmes concernant cette objection. Le premier étant que, quoi que puissent clamer les neuroscientifiques, ils ne se contentent pas d'utiliser ces termes dans leur sens technique : ils retombent constamment dans l'usage ordinaire. Selon eux, lorsque Crick (1995, p. 30), par exemple, affirme que « ce que vous voyez n'est pas ce qui est réellement là, c'est ce que votre cerveau *croit* être là », il fait intervenir toutes les connotations normalement associées à la *croyance*, non à la *croyance**. Ces dérives seraient particulièrement marquées lorsque vient le temps d'évaluer la portée explicative des théories : les neuroscientifiques prétendent offrir des théories sur la *mémoire*, la *connaissance*, l'*information*, etc., pas sur la *mémoire**, la *connaissance**, l'*information**, etc. L'autre problème envisagé par B. et H. concernant cette objection, c'est que, selon eux, les neuroscientifiques n'ont tout simplement pas encore fait le travail nécessaire pour se prévaloir de ces termes techniques : « De nouvelles règles de formation devraient être spécifiées, les conditions d'application correcte de ces phrases innovatrices devraient être spécifiées et les conséquences logiques de leur application devraient être explicitées » (p. 384), disent-ils. Nous pourrions ici répondre à B. et H. que le développement du langage ne fonctionne tout simplement pas ainsi. On a longtemps utilisé les termes psychologiques avant qu'une personne ne s'avise

de spécifier les règles comportementales de leur usage ou les conséquences logiques de leur application. L'exigence que les auteurs imposent au développement du langage et, pour ce qui nous intéresse ici, aux homonymes techniques des termes psychologiques, ne peut être satisfaite que lorsque le nouvel usage est bien établi. Imposer cette exigence à tout le langage signifierait la fin d'une large part de son évolution sémantique. Si une telle règle avait été en vigueur par le passé, des grands-parents spirituels de B. et H. se seraient élevés contre l'usage déviant du terme « atome », invoquant que seul le sens traditionnel du terme (*a-tome*, qui ne peut être divisé) possède des règles d'usage et qu'il y a des conséquences logiques bien définies à son application. Si ces règles et conséquences existent et sont connues aujourd'hui, ce n'était certes pas le cas lorsque le concept moderne d'atome était en développement. Et si on n'impose pas ces exigences ailleurs en sciences, les imposer en psychologie uniquement témoignerait selon nous d'un dualisme linguistique que nous ne croyons pas justifié. Quoi qu'il en soit, même si ces exigences sont satisfaites, le premier problème qu'ils soulèvent reste entier, pensent B. et H. Il est vrai, croyons-nous, qu'il peut y avoir une certaine confusion lors de la phase de l'établissement initial du nouveau sens technique, mais pas par la suite. Personne ne confondra aujourd'hui les deux sens de « atome » ou ne fera des inférences illégitimes du genre : les électrons sont impossibles, car ils sont posés comme étant une composante de l'atome et celui-ci, par définition, n'a pas de composantes. Si on se base sur l'histoire de la science, il nous semble y avoir de bonnes raisons de croire que l'univers sémantique des homonymes usuels et techniques se scindera progressivement, de sorte qu'un jour, les mauvaises inférences comme celles que B. et H. relèvent dans les ouvrages des neuroscientifiques contemporains ne se produiront plus. Nous devons cependant admettre que, une fois cette scission bien établie, les neuroscientifiques ne peuvent plus prétendre offrir des explications portant sur nos concepts psychologiques ordinaires. Si intéressante que puisse être une explication de la *mémoire**, ne voulions-nous pas, au départ, une explication de la *mémoire* ? Cette lacune est d'autant plus embarrassante que, dans le scénario évoqué, les termes ordinaires conservent leur usage (et donc leur sens) initial. Les neuroscientifiques pourront amoindrir un peu cet embarras s'ils parviennent à établir des explications liant la *mémoire** (concept neurologique, dont

le sens est assuré par des règles de bon usage de nature neurologique) et la *mémoire* (concept psychologique, dont le sens est assuré par des règles comportementales). Une telle explication montrerait comment la mémoire*, par exemple, la potentialisation à long terme, est impliquée dans la remémoration des connaissances par un individu. Dans le cas particulier de la mémoire, il y aura peut-être tellement de mécanismes distincts qui mériteraient d'être appelés « mémoire* » (mémoire*, mémoire**, mémoire***, etc.) que l'opportunité de créer ces sens techniques apparaîtra douteuse. Mais ce ne sera peut-être pas le cas pour d'autres concepts psychologiques. Ces considérations nous entraînent vers la dernière objection considérée par B. et H. où, cette fois, le vocabulaire psychologique ne bénéficie plus de l'immunité qui lui était accordée lors des objections précédentes.

Au chapitre 14 (p. 378), B. et H. introduisent une quatrième objection possible : que l'utilisation des prédicats psychologiques par les neurosciences constitue une innovation linguistique, justifiée pour des raisons théoriques. C'est la philosophe Patricia Churchland (1986, pp. 272-274) qui, cette fois, se porte à la défense du vocabulaire utilisé par les neuroscientifiques. Sa thèse est que le sens d'une hypothèse et le fait que celle-ci soit ou non concevable dépendent du réseau de croyances et de concepts de la personne qui aborde cette hypothèse. Une nouvelle théorie peut contribuer à modifier ce réseau et changer ainsi le sens des termes utilisés : ce qui semblait auparavant insensé cesse de l'être une fois le nouveau cadre bien intégré. Le sens original du terme, plutôt que de bénéficier d'un statut antéthéorique qui le met à l'abri de toute influence, fluctue avec la théorie. Un refus d'accepter de tels changements de sens s'apparenterait, selon Churchland, à de l'inertie sémantique. On reconnaîtra derrière cette position l'idée quiniennne d'un réseau holistique de croyances et de concepts où rien ne serait, en principe, entièrement à l'abri d'être révisé. Nous touchons ici à un point crucial, en ce que B. et H. refusent catégoriquement d'admettre l'absence d'une distinction entre énoncés analytiques et synthétiques (sur laquelle repose l'idée de Quine) ou toute conséquence que l'on pourrait en tirer (comme la continuité entre science et philosophie). « Rien, selon nous, ne pourrait être plus éloigné de la vérité », disent-ils (p. 380n.), tout en affirmant que leur livre n'est pas le bon endroit pour discuter de la chose (ils nous renvoient pour cela à Hacker, 1996, et Glock, 2003). Puisque l'ouvrage de B. et H. dit

s'adresser d'abord et avant tout à des neuroscientifiques et que ceux-ci, dans les faits, se comportent majoritairement (quoique peut-être inconsciemment) en naturalistes, il nous semble problématique de balayer le débat sous le tapis et de faire comme s'il avait été tranché en faveur de Wittgenstein. Le seul argument que les auteurs consentent à nous donner ici est que le résultat de leurs analyses justifie leur approche, ce qui, avouons-le, est un peu circulaire. De toute façon, cela leur permet de répondre à Churchland que le caractère sensé ou non d'un énoncé n'a rien à voir avec les intuitions ou les croyances de qui que ce soit, mais bel et bien avec les règles du langage, qui sont elles-mêmes déterminées par l'usage ordinaire, non théorique, des mots. Aucun fait empirique, aucune théorie scientifique, ne peut y changer quoi que ce soit. Et si une théorie scientifique s'avise de redéfinir les mots pour son propre usage, ce ne seront alors que des homonymes... ce qui nous ramène à l'objection précédente, et à la réponse de B. et H. selon laquelle les neuroscientifiques ne pourraient alors prétendre offrir des explications de nos concepts psychologiques ordinaires. L'objection de Churchland nous semble néanmoins la plus sérieuse, et l'argument que nous nous apprêtons à développer en constitue en fait un approfondissement. Mais notons d'abord que si B. et H. ont raison de dire que leur ouvrage de plus de quatre cents pages n'est pas l'endroit pour discuter du holisme quinien, alors cette courte étude critique l'est encore moins. Cependant, nous allons nous permettre la même licence que B. et H. prennent dans leur ouvrage : nous allons faire comme si celui-ci était vrai et voir ce qui s'ensuivrait. En particulier, nous montrerons qu'il y a eu en fait de tels déplacements de sens dans l'histoire des sciences, que le holisme quinien est la meilleure façon d'en rendre compte, et enfin, comme les auteurs, nous soutiendrons que cela est un argument soutenant la thèse quinienne. (Nous nous contenterons donc ici d'un match nul.)

Rappelons que l'objection précédente portait sur l'ajout d'un nouveau sens à un terme psychologique et que nous avons répondu que la contrainte imposée par B. et H. à ces ajouts de sens enfermerait la science (ou du moins le langage utilisé pour faire celle-ci) dans son état actuel. La présente objection concerne plutôt le déplacement sémantique d'un terme. Admettons donc avec Quine (et bien d'autres) que le sens de tout terme (scientifique ou non, cette distinction n'existant pas au niveau sémantique selon

l'hypothèse considérée), est déterminé par sa position dans un réseau doxastique (*web of belief*). Chaque ajout ou retrait d'une croyance dans ce réseau modifie la signification de tous les termes constituant ces croyances. On peut penser que, en général, ces changements de sens sont lents et graduels mais que, occasionnellement, avec le développement de nouvelles méthodes ou de nouveaux instruments d'observation, il se produit des changements importants qui relèguent l'ancien sens aux oubliettes. La question, évidemment, est de savoir si, de fait, on observe bien ces changements radicaux de sens. Il nous semble évident qu'ils sont fréquents. On se rappellera qu'à une certaine époque, le terme « or » référait à toute substance d'une certaine apparence jaune et brillante. Le terme était alors essentiellement phénoménal : apparence jaune et brillante. Comme tout le monde le sait aussi, Archimède a développé une méthode pour déterminer, d'une substance d'apparence jaune et brillante quelconque, s'il s'agit bel et bien de l'or. Faisant appel à la densité de la substance, qui est corrélée avec le volume d'eau déplacé lorsqu'on l'immerge, cette nouvelle procédure d'observation a radicalement changé le sens du terme « or » et de tout autre terme associé à de telles substances, pour lesquelles l'apparence devenait maintenant secondaire. Notons qu'au début, l'apparence était encore importante pour calibrer (pour ainsi dire) nos croyances liant les substances à leur densité. Éventuellement, les croyances physiques (densité, etc.) sont devenues suffisantes pour déterminer le sens de ces termes. Aujourd'hui, avec notre concept atomique d'or, on peut aisément concevoir qu'un plasma incandescent soit de l'or : l'apparence est une considération périphérique, parfois même néfaste lorsqu'elle nous amène à faire de mauvaises inférences (lançant ainsi les chercheurs sur des fausses pistes). Peut-on penser que de tels déplacements se produisent pour les termes psychologiques, dont l'essence sémantique est comportementale ? N'est-il pas possible que les croyances au sujet de comportements, croyances qui, selon l'hypothèse quiniénne acceptée pour l'argument, déterminent le sens des termes psychologiques, soient insérées dans un champ plus vaste de croyances (incluant maintenant des croyances neurologiques) et où les croyances comportementales sont d'abord utilisées, comme chez les premiers physiciens, comme contrôle des croyances neurologiques ? Ensuite ces contrôles deviendraient des preuves secondaires, utiles mais superflues. On pourrait alors

concevoir, par exemple, et contrairement à ce qu'affirment B. et H., qu'il y ait douleur sans qu'il y ait comportement de douleur. Puis, enfin, ces croyances pourraient être marginalisées dans le réseau, voire néfastes si elles s'attachent à des ressemblances comportementales superficielles ne correspondant pas aux mécanismes neurologiques qui les causent. Si cela se produit, les croyances neurologiques en viendraient à suffire pour déterminer le sens des termes psychologiques. Nous ne voyons aucune raison de croire que le fait que les croyances initiales soient de nature comportementale empêche que les croyances nouvelles soient d'une nature différente, tant qu'il existe des liens contingents unissant les deux types de croyances, par exemple : lorsqu'une personne se souvient d'un événement particulier, son hippocampe s'active. Ce cas, croyons-nous, est parfaitement analogue à celui de l'or. Il n'y a aucune raison d'immuniser la psychologie contre ces changements de sens, du moins sans adopter au préalable une attitude anti-naturaliste qui refuse à la psychologie ce qui est permis aux autres sciences.

5. Conclusion

Bien que nous ayons des doutes quant à la pertinence d'imposer le bon usage comme le font B. et H., nous croyons que tous ceux qui s'intéressent aux neurosciences devraient lire leur ouvrage pour avoir une meilleure idée de l'usage standard des verbes psychologiques dans le langage ordinaire. Les auteurs n'ont pas tort d'affirmer qu'une certaine anarchie sémantique s'est installée dans les neurosciences, et *Philosophical Foundations of Neuroscience* peut certainement contribuer à y mettre un peu d'ordre. Bien que cette anarchie sémantique caractérise en général toute nouvelle avancée de la connaissance en terre inconnue, les neuroscientifiques doivent être néanmoins conscients de l'ampleur des changements sémantiques qu'ils sont en train d'opérer et de la portée de ceux-ci sur le langage ordinaire. Un tel déplacement de sens ne peut se faire à la légère, en particulier en psychologie puisque les erreurs peuvent y être coûteuses sur le plan humain : pensons simplement aux effets qu'un mauvais diagnostic peut avoir sur un individu. Cela dit, si les neuroscientifiques sont d'accord avec les propos que nous leur avons prêtés pour répondre aux contre-objections de B. et H., ils prendront avec un grain de sel certaines admonitions de ces derniers. Quoi que les auteurs puissent reprocher à

Descartes, nous les soupçonnons d'entretenir eux-mêmes une certaine forme de dualisme, mais un dualisme linguistique plutôt qu'ontologique, où le vocabulaire psychologique bénéficie d'un statut distinct qui le place hors de portée de l'entreprise scientifique. Nous croyons, comme eux, que la philosophie doit jouer un rôle de chien de garde conceptuel mais, contrairement à eux, nous refusons d'ériger des barrières infranchissables, quitte à devoir ramener parfois au bercail quelques brebis égarées. Nous vivons une période de foisonnement scientifique et linguistique intense et (pour rester dans l'imagerie bucolique) s'il est vrai qu'il faudra éventuellement séparer le bon grain de l'ivraie, il nous semble néanmoins plus sage d'attendre de pouvoir de juger l'arbre à ses fruits.

Références

- Bennett, Max R., et Peter M. S. Hacker. *Philosophical Foundations of Neuroscience*, Blackwell Publishing, 2003.
- Blakemore, Collin. « Understanding Images in the Brain », dans H. Barlow, C. Blakemore et M. Weston-Smith, dir., *Images and Understanding*, Cambridge, Cambridge University Press, 1990.
- Blakemore, Collin. *Mechanics of the Mind*, Cambridge, Cambridge University Press, 1977.
- Churchland, Patricia. *Neurophilosophy: Towards a Unified Science of the Mind/Brain*, Cambridge, MA, MIT Press, 1986.
- Crick, Francis. *Astonishing Hypothesis: The Scientific Search for the Soul*, Londres, Touchstone, 1995.
- Damasio, Antonio R. *L'erreur de Descartes*, Odile Jacob, 1997.
- Edelman, Gerard. *Bright Air, Brilliant Fire : On the Matter of the Mind*, Harmondsworth, Penguin, 1994.
- Glock, H.-J. *Quine and Davidson on Language, Thought and Reality*, Cambridge, Cambridge University Press, 2003.
- Gregory, R. L. « The Confounded Eye », dans R. L. Gregory, et E. H. Gombrich, dir., *Illusion in Nature and Art*, Londres, Duckworth, 1973.

- Hacker, Peter M. S. *Wittgenstein's Place in Twentieth-Century Analytic Philosophy*, Oxford, Blackwell, 1996.
- Malcolm, N. « Three Forms of Memory », dans *Knowledge and Certainty*, Englewoods Cliffs, NJ, Prentice-Halls Inc., 1963.
- Ullman, Simon. « Tacit Assumptions in the Computational Study of Vision », dans A. Gorea, dir. *Representations of Vision, Trends and Tacit Assumptions in Vision Research*, Cambridge, Cambridge University Press, 1991.
- Wittgenstein, L. *Investigations philosophiques*, Paris, Gallimard, 1961.
- Young, J. Z. *Programs of the Brain*, Oxford, Oxford University Press, 1978.